DOCUMENT RESUME

ED 358 109 TM 019 875

AUTHOR Freeman, Jared T.; Guillen, Julio
TITLE Formative Evaluation of a Generic Decision Aid for Classroom Use.
PUB DATE Apr 93
NOTE 7p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Biology; *Classification; Comparative Analysis; Computer Uses in Education; Databases; Decision Making; *Formative Evaluation; Higher Education; Information Retrieval; *Search Strategies; Statistics; *Undergraduate Students; *User Needs (Information)
IDENTIFIERS Cognitive Flexibility; *Decision Aids; *XPT EASE

ABSTRACT

Results of a formative evaluation of a decision aid for students of taxonomic domains such as statistics or biology are reported. The tool, XPT-EASE, is designed to allow a student to search a taxonomy by traversing its branches in an arbitrary order, presumably the order simplest for the student, rather than by starting from the root node and proceeding from one connected branch to the next. XPT-EASE is a generic shell for decision aids. For this study, it was equipped with a database concerning statistical methods. Eight undergraduates in a statistics course participated, half using a flexible version of XPT-EASE and half using a static version. In the flexible version, the attributes were presented in a single menu and their order was randomized on every presentation, while in the static version, only one attribute was presented at a time. The flexible version pared the list of candidate techniques as the students selected attribute values, but in the static version, the list was not trimmed until all selections were made. Results indicate that the flexible version boosts speed and accuracy with which subjects identify statistical techniques relative to the static version, which presents queries in a set order as do traditional decision aids. (SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

# FORMATIVE EVALUATION OF A GENERIC DECISION AID FOR CLASSROOM USE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Jared T. Freeman, Teachers College, Columbia University, NY
Julio Guillen, Jersey City State College, NJ

Presented at the 1993 annual meeting of
The American Educational Research Association,
Atlanta, Georgia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JARED FREEMAN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED358109

## ABSTRACT

This paper reports the results of a formative evaluation of a decision aid for students of taxonomic domains such as statistics or biology. The tool, called XPT_EASE, is designed to allow a student to search a taxonomy by traversing its branches in an arbitrary order, presumably the order that is simplest for her, rather than by starting from the root node and proceeding from one connected branch to the next. XPT_EASE is a generic shell for decision aids. For this study, it was equipped with a database concerning statistical methods. The study indicated that the flexible search scheme boosted speed and accuracy with which subjects identified statistical techniques used to solve word problems, relative to a version of the tool that presented subjects with queries in a set order, as traditional decision aids do.

## INTRODUCTION

When educators discuss software, we generally consider two types of programs: educational packages, such as simulations and drill-and-practice programs, and productivity software, such as word processors and spreadsheets. A third type of program has received little attention, but it may play an important role in school and college classrooms as laptop and palmtop computers become more common. It is the computerized decision aid, a type of program that a student will use as a professional does in the working world, to help identify relevant features of a problem and select an appropriate solution strategy or algorithm.

Decision aids are not new to classrooms. They are commonly found in textbooks on statistics, biology and the physical sciences, often in the form of a taxonomy. For example, a student may use a taxonomy of the animal kingdom to identify the name of a creature (or class of animals) given some characteristics, or the characteristics of a creature given its name. We are particularly interested in decision aids for students of statistics, in part because authors have created a wealth of printed decision aids and computerized ones that do not serve students well.

Printed taxonomies of statistical methods (such as those by Tabachnick and Fidell, 1989; Andrews, et al., 1981) list a range of problem characteristics along their branches (e.g., type of data: nominal, ordinal, interval, ratio; number of variables) and statistical methods at the leaves (e.g., z-test, t-test, ANOVA). Such taxonomies make poor decision trees for two reasons. First, a student can search a tree correctly only if she starts at the root or the leaves. To start amidst the central branches is to skip potentially important decisions points. Second, the only student who can complete a search of a taxonomic tree is one who understands the meaning of labels on all of the branches and leaves of the along the search path. These are serious limitations. They can be summarized as follows: printed taxonomies do not afford individualized search strategies, nor do they compensate for poor understanding of terms in the domain.

One method of making printed decision trees more useful is to convert them to tables, with parameters on the axes and solutions in the cells. Because the student can "search" the matrix beginning on either axis with any item, the tool affords customized searches. For example, a simple matrix such as the following allows the student to begin the search for a technique by identifying either the number of samples or the character of the data.

TM019875

| | One sample | Two samples |
|---|---|---|
| Discrete data | Chi-square test of goodness of fit | Chi-square test of independence |
| Continuous data | One-sample z-test | T-test |

Caption: A table of several statistical techniques differentiated along two dimensions.

However, the matrix format is a poor choice for domains with more than a few dimensions. As dimensions are added to a matrix, the designer must either increase the number of cells dramatically, or split the matrix into submatrices. Statistics is the sort of complex, multi-dimensional domain that is poorly represented by tables.

The second method of resolving the problems with printed taxonomies is to automate them. Balian's Select-Stat (1987), Timko and Downie's Statistics on Software (1992) and The Idea Works (1992) have done just this. Their aids for selecting statistical techniques require the subject to answer as many as 15 questions to isolate an appropriate statistical technique. Two of the programs address the problem of unfamiliar terminology by providing definitions on demand, through a help function.

All facilitate search by presenting sequences of questions that presumably minimize the length of any given search. However, the order of those questions is fixed along each search path. Thus, it is possible that a student will encounter a question that is perplexing (despite the definitions offered in an on-line glossary) early in a session, when the search space is least constrained. She may become frustrated, make incorrect decisions, and arrive at an incorrect solution. Experts often strategize specifically to avoid this situation, by answering the easiest questions first (Riedl, et al., 1991). However, automated decision aids of the type just described foil both expert strategists and the very students the designers intend to help, insofar as the tools prohibit the user from answering the easiest questions first. In this respect, existing automated taxonomic search aids are no improvement over printed taxonomies.

## THE DECISION AID: XPT_EASE

We have developed and conducted formative evaluation of a domain-independent tool that affords individualized search of a taxonomy, and provides definitions on demand. It is called XPT_EASE. The program functions on IBM-compatible computers and is written in Prolog, to capitalize on the backtracking engine underlying that language. It performs two functions:

- It can identify an item (such as a bird) if the user names its attributes (it has wings and feathers); and
- It can identify the attributes of an item the user names.

These are not unusual functions, but the way the XPT_EASE performs them is novel among statistical decision aids we have reviewed. If a student opts for the first of the two options above, she is presented with a screen consisting of two windows. The top window lists all "*items*" in the database. In the study described below, the items were ten statistical techniques such as Pearson r, ANOVA, and simple regression. The lower window is empty as the session begins. A menu overlaying the lower window presents several questions, or "*attribute queries*", from which the student can select the easiest or most salient. For example, in the statistics study, below, one attribute query is "Are the data nominal or ratio?" When the student selects any one of these queries, the question is displayed in the lower window, and the menu contents change to list possible answers (or "*attribute values*") to the query, such as "nominal data" and "ratio data". When the student selects a value, such as "ratio data," the system responds in three ways:

- In the upper window, it deletes all items that do not have the selected attribute value and all others selected previously. Given a single attribute value of ratio data, in the statistics version of XPT_EASE, below, the system would eliminate from the upper window chi-square tests for independence and goodness of fit, leaving only methods that

employ ratio data. Thus, the system maintains a visible and minimal representation of the candidate items.

- In the lower window, the system displays the attribute query and the selected attribute value. Thus, it maintains a public trace of the student's reasoning.
- On the menu, the system displays all the remaining attribute queries, less those for which values are the same for all remaining items in the upper window. For example, if the reduced list of statistical techniques included only those that required two variables, the query concerning number of variables would be removed from the menu. Thus, XPT_EASE intelligently minimizes the number of decisions the student must make to arrive at a solution, and the number of attributes the student must consider at a given time.

As the student continues to select attributes and their values, the system trims the list of candidate items displayed in the upper window, expands the trace of previously selected attributes and values in the lower window, and shortens the list of remaining attribute queries in the menu, until there are no more queries that distinguish between remaining items. It then announces the solution(s).

The user can select any attribute query from the query menu. Queries need not be answered in a set order, as they must in similar automated decision aids and as is implicit in paper taxonomies. Pressing the standard help key (F1) at this or any time displays a definition of the current menu item.

The system also performs a second function, as noted above; it can list the attributes and attribute values of a given technique. The subject simply selects one of the list of items (e.g., statistical techniques). The system then displays that item in its upper window, and presents in the lower window each attribute query and its value for that item. This function is not of interest to us in the present research.

For research and testing purposes, XPT_EASE keeps a time-stamped trace of every menu choice and help query by the user.

## A FORMATIVE EVALUATION OF XPT_EASE

XPT_EASE allows a user to construct her own search path by identifying attributes and their values in any order. Thus, the system presents the user with a "*flexible taxonomy*" which we find intuitively appealing. However, it could be argued that the printed "*static taxonomy*" and the rigidly ordered queries of automated statistics decision aids serve an important function: they may encourage the student to learn predetermined and optimal search paths. Those paths are optimal presumably because they capitalize on the fundamental, hierarchical structure of the domain.

We set out to examine the efficacy of flexible and static taxonomies during the formative evaluation of XPT_EASE. Specifically, we tested the following hypotheses:

- Given the opportunity, students will initiate searches at different branches of a taxonomy.
- Individualized search will produce faster and more accurate performance.

### Subjects

Eight undergraduates students participated in the study. All were enrolled in an introductory, summer statistics course at Jersey City State College. Most were underpriveleged, inner city youth, and for some, English was a second language. Subjects were paired by ability and divided into two groups. One group used a "flexible version" of XPT_EASE and the other a "static version." (Both versions are described below).

### Materials

The two versions of the program were identical in two respects. Both used the same database of statistical techniques and their attributes. That database is represented in the following table:

| 1. Do you want to make inferences about samples or about variable relations? | 2. Are the data nominal or ratio? | 3. How many samples are there and how are they related? | 4. How many variables are there? | Statistical technique |
|---|---|---|---|---|
| Samples | Ratio | Two dependent samples | One dependent variable | T-test for two dependent samples (2 groups) |
| Samples | Ratio | Three or more independent samples | One dependent variable | One-way ANOVA for 3 or more samples |
| Samples | Ratio | Two independent samples | One dependent variable | Z-test for two independent samples |
| Samples | Ratio | Two independent samples | One dependent variable | T-test for two independent samples |
| Samples | Ratio | One sample | One dependent variable | One-sample T-test |
| Samples | Ratio | One sample | One dependent variable | One-sample Z-test |
| Samples | Nominal | One sample | One variable | Chi-square goodness-of-fit test |
| Relations among variables | Ratio | One sample | Two variables | Simple regression |
| Relations among variables | Ratio | One sample | Two variables | Pearson r |
| Relations among variables | Nominal | One sample | Two variables | Chi-square test for independence |

Caption: Database for selecting statistical techniques used in a formative evaluation of XPT_EASE.1 Notes: The T-test for two dependent groups includes the paired-samples T-test and pre- post usage. The query concerning number of variables concerns only those variables that can properly be called dependent. The dependent/indpendent distinction does not exist for statistics that describe the relations between variables.

Second, both versions implemented all of the functions and features described previously with the following exceptions:

- In the flexible version, all as-yet-unselected attributes were presented in a single menu and their order was randomized on every presentation to minimize the implication of a recommended search path. In the static version, only one attribute was presented at a time, the order of presentation was constant and the attribute queries were numbered. This was intended to reflect the static root-to-leaf search paths implied by printed taxonomies and traditional decision aids.
- In the flexible version, the displayed list of candidate techniques was pared as the student selected attribute values. In the static version, this list was not trimmed until the student had made all necessary selections of attributes and values.

---

[1] To simplify the system for statistical novices, the database did not query users concerning sample size. Thus, identical paths led to both the T- and Z-tests. The attribute of sample size could be added to the database, in which case the user who indicated small sample size would be advised to use a T-test; a large sample (n=30) would produce the recommendation to use a Z-test. For very large samples, either test would be appropriate.

## Method

Each group of subjects was independently introduced to its version of the system and given the opportunity to practice selecting techniques to solve statistics word problems. The introductory session lasted 20 minutes. All subjects were then given a set of 44 word problems selected from standard statistical texts and were asked to identify the best technique(s) to use for each problem. Subjects were told that the test was speeded (only 45 minutes were allocated for completion) and that guessing was penalized. Students completed as many problems as possible during the time allotted. The system recorded a time-stamped trace of every action by each student. At the end of the session, a questionnaire concerning the quality of the system was administered.

## Results:

Subjects using the flexible version of XPT_EASE took advantage of the opportunity to individualize their search paths (H1). The most common step subjects took to begin a search was to answer the query "How many samples are there and how are they related?" However, two of the subjects selected another attribute query first more that half the time, and all started with a different query at least one-third of the time. The query least frequently chosen to begin a search concerned the quality of the data, nominal or ratio, confirming anecdotal evidence that data types are particularly confusing to novices. Overall, there was a significant interaction between subjects and the frequency with which each attribute query was the first query in a search ($F = 6.306$, $p = .008$). Three of the four subjects started at least one search with each of the four attribute queries. The fourth subject started each search with one of three attribute queries. In contrast, subjects in the static condition were forced to start searches by addressing the same attribute query, and they answered remaining queries in a set order. The finding that choice of starting point varied by between and within subjects supported the first hypothesis.

Subjects using the flexible version answered a more questions correctly (mean = 20.5, or 77% of all questions answered) than subjects using the static version (mean = 8, or 49% of all questions answered). Because of the low n (four in each group) a nonparametric method was employed to test the significance of this difference. The Mann-Whitney U was significant at alpha = 0.05. (The difference was marginally significant when tested using a parametric method: $t = 2.338$, $p = 0.058$).

Subjects using the flexible version were also faster, answering more questions overall (mean = 26.5) than subjects using the static version (mean = 17.5). This difference was significant at alpha = 0.05 using the Mann-Whitney U test. (A t-test was not significant: $t = 1.4$, $p = 0.28$). The findings concerning accuracy and speed supported the second hypothesis.

It was not possible to determine if subjects in the flexible condition most often addressed the easy queries first because the planned index of ease, namely the frequency of reference to help for a particular query, was maldistributed between subjects. Two subjects did not use help at all, two used it less than five times, one a dozen times, and the remainder used it 23 to 44 times.

Finally, the debriefing revealed that subjects were enthusiastic about XPT_EASE, regardless of the version they had used. Many wanted to make use of the tool on their final exams, and they were allowed to do so. Some asked that the tool be integrated with statistics computing packages, such as MINITAB. Finally, there was general concern that the definitions provided by the help function be improved. All definitions contained an abstract explanation follwed by an example.

## DISCUSSION

These results supported the proposed hypotheses. Flexibility in selecting a search path boosted the speed with which students generated answers using this decision aid, and the quality of those answers. Contrary to the assumption made by designers of existing decision aids, a fixed path through a search space is not necessarily the best for any given student.

XPT_EASE implements an effective model for automated decison aids. However, there is a serious limitation to this tool of which designers should be aware. First, the search logic of XPT_EASE requires that every item (such as ANOVA) have at least one value for each attribute

of every item in the database. Thus, if the T-test has the attribute of data type (and the specific value, ratio) then ANOVA must also be defined such that it has a value for that attribute. In the worst case, an attribute of one item (say, color of feathers for the item, birds) may be irrelevant to another item (for example, fish). Also problematic is the case where the number of attributes in the database is so great that the user faces an overwhelming number of attribute queries. In these domains, designers should consider applying the search logic of XPT_EASE to the intersection of database attributes, if the common attribute queries are few in number (say three to ten), and a traditional AI approach to the remainder of the database. Thus, users would benefit from the XPT_EASE approach during the initial stages of their search before resorting to the more cumbersome strategy enforced by commercial products such as the ones described above.

Further research concerning XPT_EASE is in order. First, it would be useful to know if domain novices traverse the attribute query list from the easiest question to the most difficult. This is a strategy that experts claim to employ, and it has considerable surface validity. The question might be examined by measuring the "easiness" of questions either by the number of times each student invokes the help function to explain that item, or by each subject's confidence in their knowledge of the item on a pre-test survey. If domain novices do answer the easy questions first, then this putatively expert strategy need not be taught to them. If novices do not use use strategy, perhaps they would benefit by learning it.

Second, the manner in which search paths vary by problem type is inherently interesting. If XPT_EASE was modified to record problem numbers, it would be possible to trace changes in the search paths of students as a function of the type or structure of statistics word problems.

Third, the current implementation of XPT_EASE as a statistics aid might be improved. The attributes and attribute values might be clarified. The help text could be structured to provide, in addition to an abstract definition and example, pointers to related terms and comments that distinguish similar terms. In addition, the help database might be made extensible, so that students can make notes that clarify difficult concepts. These annotations would be useful feedback for the database developers.

In sum, the results presented here can guide the designers of decision aids for the classroom and raise new questions concerning the manner in which students search taxonomic domains.

## NOTES

This work was partially funded by a grant from the State of New Jersey, administered by Jersey City State College.

## BIBLIOGRAPHY

Andrews, Frank M., Laura Klem, Terrence N. Davidson, Patrick M. O'Malley and Willard L. Rogers. (1981). A Guide for Selecting Statistical Techniques for Analyzing Social Science Data (Second Edition). Ann Arbor, MI: Institute for Social Research, The University of Michigan

Balian, Edward S. (1987). Select-Stat. Software for Research.

Riedl, T.R., Weitzenfeld, J.S., Freeman, J.T., Klein, G.A., and Musa, J. (1991). What we have learned about software engineering expertise. In J.E. Tomayko (Ed.), Software Engineering Education: SEI Conference 1991. NY: Springer-Verlag.

Tabachnick, Barbara G. and Fidell, Linda S. (1989). Using Multivariate Statistics. NY: Harper Collins.

The Idea Works. (1992). Statistical Navigator Professional. Columbia, MO.

Timko, John and Jon Downie. (1992). Statistics on Software. Orange, CA: Statistics for Management.